

For Completion by NILS-RSU only Project Number:



NILS-RSU

NORTHERN IRELAND LONGITUDINAL STUDY
RESEARCH SUPPORT UNIT

NILS/NIMS APPLICATION FORM

The Northern Ireland Statistics and Research Agency (NISRA) is responsible for the Northern Ireland Longitudinal Study (NILS) and the Northern Ireland Mortality Study (NIMS).

Please refer to the NILS-RSU website for the RAG Applications Criteria which are used to assess a research proposal. [See link here](#)

PART A: USER AND STUDY DETAILS

SECTION A1 APPLICATION DETAILS

Study Title

This should be a brief informative summary of the research proposal.

Community Segregation in Northern Ireland

Proposed Start Date (DD/MM/YY)

Proposed End Date (DD/MM/YY)

Please ensure that you propose a realistic end date. This can only be amended at a later stage in exceptional circumstances.

Contact Details of the Chief Researcher

Title	
Name	
Position	
Organisation	
Address	
Telephone	
Email Address	
Date of last Safe Researcher Training (SRT)	

Contact Details of Other Applicant(s)

Please provide details of all other researchers who will be using the data.

Name	Organisation	Email Address	Date of last SRT



I agree to protect the confidentiality of the NILS/NIMS data.



I will follow all the 'Dos and Don'ts' as published by the NILS team.



I will read and sign the NILS Licence Agreement, NILS Disclosure Control Policy and NILS Security Policy upon approval of this application.

Signature:

Date:

(electronic signatures accepted here)

SECTION A2 BACKGROUND AND AIMS OF THE STUDY**Project Summary/Abstract**

This will be published on the NILS-RSU website if the research proposal is approved (circa 250 words). [See current projects](#)

I aim to make use of information about residential area choices and the timing of moves to uncover the preferences for housing, segregation and other dynamically-evolving area attributes in Northern Ireland (NI). The 2001 Census indicates that roughly 42% of Catholics and 39% of Protestants live in areas (1 km grid squares) where over 90% of the people are of the same community background as them (Shuttleworth and Lloyd, 2009). Many papers have considered the degree of segregation and what the consequences are but few try to establish what are the causes and dynamics that perpetuate it and what households' preferences are actually over.

For NI specifically it is possible that violence (or the fear of violence) goes a long way to explaining segregation and therefore looking at the effects of crime and violence on location decisions will be important. Education and income are likely to also play a role - with more mixed areas tending to consist of better educated and wealthier households - part of what I seek to establish is how large a role.

This study is rooted in both the Tiebout sorting literature and the more general topic of segregation. I will quantify preferences by estimating the marginal willingness to pay for area attributes by applying a dynamic discrete choice model for housing and neighborhood attributes.

Relevance of Research Proposal to NILS/NIMS

Please provide evidence of the longitudinal aspect to the research [\[see guidance here\]](#) and explain how the NILS/NIMS data is uniquely placed to contribute to the study.

The project will use information about residential area choices of households and the timing of moves (across years) to uncover moving costs and preferences for housing and area attributes.

There are three main reasons why locational decisions are inherently dynamic: large costs (both financial and psychological) make moves relatively rare; households' tastes and needs change over their life-cycles; and evolving local amenities and housing prices give neighbourhoods a dynamic character. The longitudinal/dynamic aspect is also important as it can be shown that if only considered in a static setting any estimates of household preferences will be biased depending on the time series properties of the factors being considered (see Bayer et al. 2011). For instance, consider the community background composition of an area, this will be highly persistent over time (i.e. if an area is primarily Catholic in one year then it is highly likely it will also be primarily Catholic in the next year), given this fact a static discrete choice model would overstate the importance of community background to individuals as it would be ignoring the fact that individuals know that this characteristic of the area will persist over time.

Northern Ireland provides an interesting case of very split community relations, from which it is incredibly interesting to try to measure preferences. My aim is to quantify

preferences by calculating the monetary value (i.e. marginal willingness to pay) that individuals with different characteristics place on the non-market features of those areas. The NILS is uniquely placed to contribute to the study as it contains rich information about households characteristics (e.g. community background), property attributes, (can be linked to) area attributes (e.g. crime rates) and crucially it contains the household moving decisions that are required to build a model of this type.

Please provide a brief summary of what is known to date about your chosen topic and how the proposed research would contribute.

Tiebout (1956) suggests that consumers reveal preferences for local public goods by the residential locations they choose. Since then there has been a large (both theoretical and empirical) literature on residential sorting and this has been extended to look at non-market amenities, with Bayer et al. (2007) finding that households have strong preferences to self-segregate on the basis of education and especially race.

Within sociology and geography there is a large literature on segregation, with the most attention paid to developing various constructions of segregation measurement. In economics generally research is US based and looks at race, such as Cutler and Glaeser (1997) who find that black people in more segregated areas have significantly worse outcomes than blacks in less segregated areas. Other studies, such as Echenique and Fryer (2007) and Cutler et al. (2008), stress the importance of neighborhood effects and social networks suggesting that segregated neighborhoods could contribute to racial gaps in a variety of socioeconomic outcomes. The consequences of segregation are well studied but what is lacking in the literature is an understanding of the dynamics behind why segregation is occurring. An issue which is often discussed but needs more attention is whether segregation is largely an effect of income differences rather than race (or religion).

Getting a better grasp on some of the causes and dynamics of segregation and by establishing household preferences is an understudied area in the literature on segregation and any results should offer policy implications. The vast majority of papers on segregation look at the US however residential segregation is a universal feature of almost every city and country in the world and goes beyond the dimension of just race. Segregation also occurs by factors such as income and increasingly important is segregation by religion. My study would offer evidence from a different country and on the dimension of religion rather than race.

Relevance of Research Proposal to Health & Social Care

Please indicate how the research proposal relates to Health & Social Care research [\[see guidance here\]](#).

Where people live has a huge impact on their social well-being and quality of life. Homes where people feel safe and that are surrounded by amenities they value help people to live healthier and happier lives. This fact is recognized by the Office of National Statistics who include *where we live* as one of the important domains in their measurement of *National Well-being* (see ONS, 2012). It is also recognized by the The Marmot Review (2010) which links social cohesion and where people live with health and social inequality, giving the example of whether you live in a safe neighbourhood without fear of crime as something that will affect your well-being.

For Northern Ireland in particular where people live goes beyond just social well-being but has been linked with both mental (e.g. O'Reilly and Stevenson, 2003) and physical well-being (e.g. Mueller et al. 2013).

My research aims, by analyzing the locational choices of households, to uncover the preferences that individuals place on area characteristics and on segregation, which should contribute to the understanding of what affects wellbeing and by how much.

Please indicate how the research may support the development/delivery of public policy, public service delivery or serve the public good. Please reference any local or national policies where possible.

Two fairly recent policy documents from the Office of the First Minister and Deputy First Minister, *A Shared Future: Improving Relations in Northern Ireland* (2005) and the *Programme For Cohesion, Sharing And Integration* (2010), both establish segregation as a major problem for social cohesion that leads to conflict. Even if you ignore the huge social cost of conflict, a report by Deloitte (2007) put the cost of segregation in NI through the duplication of public services alone at (an upper limit of) £1.5 billion a year.

If my work can help to provide a better understanding of what influences location decisions and segregation in Northern Ireland and what it is that different types of households value this can aid policymakers in designing more effective interventions.

Prior Experience of Researcher

Please provide detail of your experience or understanding of using longitudinal data and of handling potentially disclosive microdata.

During my studies (BSc Econometrics & Mathematical Economics, MSc Economics, and MRes Economics) I have completed many microeconometrics courses which have had substantial sections devoted to longitudinal data.

I have also worked as a Research Assistant on various longitudinal projects, including one using a restricted version of the British Household Panel Survey (BHPS) and another using confidential Norwegian municipal data.

To familiarize myself with the NILS I have read the *NILS Cohort Profile* (2011) and *NILS Working Paper 1.0 – An Introduction to NILS* (2010) as well as some papers that use the

NILS (e.g. Shuttleworth and Barr, 2011). These readings have given me not only a good understanding of what is contained in the NILS but also about the procedures in place to protect the confidential nature of the data.

Planned Publication and Dissemination of Findings

Please provide detail on the relevant stakeholders for the research.

<input type="checkbox"/>	Health bodies such as Department of Health, Social Services and Public Safety, Public Health Agency, Health and Social Care Board, Department of Health (England); please specify:
<input type="checkbox"/>	Other Government Department; please specify:
<input checked="" type="checkbox"/>	Research Council, Charitable Foundation; please specify: <i>ESRC (funding my PhD)</i> <i>Northern Ireland Community Relations Council (NICRC)</i>
<input type="checkbox"/>	NISRA; please specify:
<input checked="" type="checkbox"/>	Other; please specify: <i>Research departments - e.g. UCL</i>

Indicate how you plan to publicise and disseminate the findings of the study.

<input checked="" type="checkbox"/>	Refereed journals <i>An Economics journal - e.g. Journal of Public Economics, Journal of Political Economy, Journal of Urban Economics etc.</i>
<input checked="" type="checkbox"/>	Report for Government Department, Research Councils or other sponsors <i>Any findings will be reported to the NICRC</i> <i>The ESRC are informed of research findings</i>
<input type="checkbox"/>	Internal NISRA reports
<input checked="" type="checkbox"/>	Presentations (conferences, workshops, seminars) <i>Seminars - e.g. NICRC, universities (UCL and others)</i> <i>Conferences - e.g. ESRC London DTC conference</i>
<input type="checkbox"/>	Other; please specify:

You will be required to periodically provide updates on any policy relevance, impact or dissemination activities. This will be throughout the life of the project and for a period afterwards.

PART B: SPECIFYING DATA EXTRACT

SECTION B1 STUDY POPULATION

Dataset to be Used (Please tick). NILS NIMS

Study Population

Please provide a description of the population selected for the study specifying as much detail as possible. Please consider gender, age group, timeframe and the link to census data.

For example, all male NILS members aged between 15-44 enumerated at the 2001 census.

Full sample:

All households containing NILS members enumerated in the 2001 Census (those members living in community housing will be excluded and so will student households).

Homeowner subsample:

Households contained in the full sample which are homeowners (i.e. renters are excluded)

Vital events data:

Part of what I seek to establish is what types of households are more likely to move. The BIRTHSSTATS and DEATHSSTATS data will enable me to establish whether a household has undergone any change in circumstance or composition which may have required a move (e.g. a childless couple in 2001 move home in 2004 – this may be due to the birth of a child and the requirement for a larger home, or an elderly couple in 2001 move home in 2009 – this may be due to the death of one partner and the requirement to live in a single storey home).

Estimated Sample Size

Please include details of the size of the sample and any subgroups that may be the focus of any analysis. Alternatively provide power calculations [\[see guidance here\]](#).

Full sample:

- In Northern Ireland there were c.628,400 households in 2001 (this excludes communal establishments)
- Given the NILS sample (c.28% individuals in population) and there being approximately 2 people per household this would give c.50% or **c.314,200 households in the full sample**
- The **event population is an average of c.25,136 per year**, based on the fact that roughly 8% of households move each year. However please note that I will also gain information from all the other households that choose not to move in a given year.

Homeowner subsample:

- Roughly 70% of households are homeowners in 2001 (when communal establishments are excluded), which gives **c.219,940 households in the homeowner subsample**
- The number of property sales has varied considerably by year – for instance a high of 41,387 in 2006 to a low of 10,790 in 2010 (LPS, 2014). Given that c.50% of all households are in the full sample I would expect the **event population in this subsample to be between c.5,395 and c.20,694 per year**. Again please note that I will also gain information from the homeowners that choose not to move.

Super Output Areas:

- Part of this research will involve looking at Super Output Areas (890), which would give an average of c.200 households per year, that would be averaged over, in each year.

NILS Core datasets**Census data**

1991	<input type="checkbox"/>
2001	<input checked="" type="checkbox"/>
2011	<input checked="" type="checkbox"/>

Vital events data

Births of NILS Member	<input checked="" type="checkbox"/>	from 2001	to 2012
Births to NILS Mothers	<input checked="" type="checkbox"/>	from 2001	to 2012
Births to NILS Fathers	<input checked="" type="checkbox"/>	from 2001	to 2012
Deaths of NILS Member	<input checked="" type="checkbox"/>	from 2001	to 2012

Event dates**Other NILS data**

Internal Migration	<input checked="" type="checkbox"/>	from 2001	to 2012
Immigration/Re-entrants	<input checked="" type="checkbox"/>	from 2001	to 2012
Emigration	<input checked="" type="checkbox"/>	from 2001	to 2012

Variables Required

Please complete this section using the NILS Data Dictionary and, if necessary, in consultation with the NILS-RSU. Please add more rows as necessary. Please note this dataset can only be updated at a later stage through a project modification form. [Weblink to data dictionary](#)

Table Name	Variable Name	Justification for Inclusion (necessary for all Restricted variables)
CoreNILSData		This dataset will be used to link the rest together.
CoreNILSData	NILSID	
CoreNILSData	SOURCE	
CoreNILSData	GENDER	
CoreNILSData	STATUSHISTORY_FULL	
CoreNILSData	STATUSHIST	
CoreNILSData	CURRENT_ADDRESS_SOA	
CoreNILSData	CURRENT_ADDRESS_XUPRN	Needed to link the properties data.
CoreNILSData	DODMM	
CoreNILSData	DODYEAR	
CENSUSHH_2001		This dataset is where the most of the information on households will come from.

CENSUSHH_2001	XUPRN	Required to link each household to the house they are living in.
CENSUSHH_2001	CENSUSHID0	
CENSUSHH_2001	SOAENUM	
CENSUSHH_2001	HHOCCSTATH0	
CENSUSHH_2001	HHSIZEH0	
CENSUSHH_2001	ACCTYPEH0	
CENSUSHH_2001	TENUREH0	
CENSUSHH_2001	SELDCONTH0	
CENSUSHH_2001	HH17PLS_COUNTH0	
CENSUSHH_2001	HHWRKG_COUNTH0	
CENSUSHH_2001	HHPEN_COUNTH0	
CENSUSHH_2001	HHADULST_COUNTH0	
CENSUSHH_2001	HHDEPCHLD_COUNTH0	
CENSUSHH_2001	HHSTUHOME_COUNTH0	
CENSUSHH_2001	HHADEMP_COUNTH0	
CENSUSHH_2001	HHCOMPH0	
CENSUSHH_2001	HHFAMTYPEH0	
CENSUSHH_2001	HHPENSH0	
CENSUSHH_2001	HHADCHLDSTRH0	
CENSUSHH_2001	HHDEPCHLDH0	
CENSUSHH_2001	HHCARERSH0	
CENSUSHH_2001	HLLTIH0	
CENSUSHH_2001	STUAWAYH0	
CENSUSHH_2001	HHCOMBACKH0	
CENSUSHH_2001	HHETHRELH0	
CENSUSHH_2001	HHETHSTRH0	
CENSUSHH_2001	HHWKRSTRANSH0	
CENSUSHH_2001	HRP_SEXH0	
CENSUSHH_2001	HRP_AGEH0	
CENSUSHH_2001	HRP_MARSTATH0	
CENSUSHH_2001	HRP_COMMBACKH0	
CENSUSHH_2001	HRP_RELH0	
CENSUSHH_2001	HRP_COBH0	
CENSUSHH_2001	HRP_ETHGRPH0	
CENSUSHH_2001	HRP_EDHLQH0	
CENSUSHH_2001	HRP_EACTH0	
CENSUSHH_2001	HRP_OCCH0	
CENSUSHH_2001	HRP_INDH0	
CENSUSHH_2001	HRP_NSSH0	
CENSUSHH_2001	HRP_SOCGRDH0	
CENSUSHH_2001	HHFAM_COUNTH0	
CENSUSHH_2001	HHCARERS_COUNTH0	
CENSUSHH_2001	HLLTI_COUNTH0	
CENSUSHH_2001	OWNERSHIPH0	I plan to differentiate between owners and renters so I need to know who is a renter or an owner.
CENSUSHH_2001	NS_DEP_EMPH0	
CENSUSHH_2001	NS_DEP_EDUH0	
CENSUSHH_2001	NS_DEP_HEAH0	
CENSUSHH_2001	NS_DEP_HOUSH0	

CENSUSHH_2001	NS_DEP_TENH0	
CENSUSHH_2011		This dataset will be used to see what household details have changed since the 2001 Census.
CENSUSHH_2011	CENSUSHID1	
CENSUSHH_2011	ADEMH1	
CENSUSHH_2011	ADTH1	
CENSUSHH_2011	AHCH1	
CENSUSHH_2011	AHTH1	
CENSUSHH_2011	CARSNOH1	
CENSUSHH_2011	CECTMCEWSH1	
CENSUSHH_2011	HRPCOBH1	
CENSUSHH_2011	CRSH1	
CENSUSHH_2011	DEPEDH1	
CENSUSHH_2011	DEPEMH1	
CENSUSHH_2011	DEPHDH1	
CENSUSHH_2011	DEPHSH1	
CENSUSHH_2011	DEPRIVEDH1	
CENSUSHH_2011	DEPTNH1	
CENSUSHH_2011	DPCH1	
CENSUSHH_2011	EILAH1	
CENSUSHH_2011	ETHH1	Useful to know the ethnicity of the household to be able to compute ethnicity of the area.
CENSUSHH_2011	FAMH1	
CENSUSHH_2011	HHCH1	
CENSUSHH_2011	HHLSH1	
CENSUSHH_2011	IHC1H1	
CENSUSHH_2011	IHC2H1	
CENSUSHH_2011	ILAH1	
CENSUSHH_2011	ILLADULTH1	
CENSUSHH_2011	ILLH1	
CENSUSHH_2011	ILLLITTLEH1	
CENSUSHH_2011	ILLLOTH1	
CENSUSHH_2011	MEIGH1	Useful to know the ethnicity of the household to be able to compute ethnicity of the area.
CENSUSHH_2011	HEACONAH1	
CENSUSHH_2011	HEACONH1	
CENSUSHH_2011	NSSH1	
CENSUSHH_2011	NSTAH1	
CENSUSHH_2011	P17H1	
CENSUSHH_2011	PENH1	
CENSUSHH_2011	PPROOMH1	
CENSUSHH_2011	ROOMREQH1	
CENSUSHH_2011	ROOMSH1	
CENSUSHH_2011	SELFCONH1	
CENSUSHH_2011	SIZH1	

CENSUSHH_2011	STAH1	
CENSUSHH_2011	TENH1	
CENSUSHH_2011	TENUREH1	
CENSUSHH_2011	TYPACCOMH1	
CENSUSHH_2011	USHC1H1	
CENSUSHH_2011	USHC2H1	
CENSUSHH_2011	NORDH1	
CENSUSHH_2011	TENDH1	
CENSUSHH_2011	HHSDH1	
CENSUSHH_2011	EA_SOA_CODE_2011H1	
CENSUSHH_2011	TENURE_IMPH1	
CENSUSHH_2011	XUPRN	Needed to link the properties data.
PROPERTIES		This dataset is where details on the properties will come from.
PROPERTIES	XUPRN	Needed to link the properties to families in each year.
PROPERTIES	SETTLEMENTBAND	
PROPERTIES	CV_NON_EX	Needed to calculate valuation of each property.
PROPERTIES	PRIMARY_CLASS	
PROPERTIES	SUB_CLASS	
PROPERTIES	TYPE	
PROPERTIES	HAB_SPACE	
PROPERTIES	ANC_SPACE	
PROPERTIES	HAB_ROOMS	
PROPERTIES	TOTAL_BED	
PROPERTIES	BATHS	
PROPERTIES	HALF_BATHS	
PROPERTIES	HEATING	
PROPERTIES	YEAR_BUILT	
PROPERTIES	STOREYS	
PROPERTIES	FLOOR	
PROPERTIES	PARKING	
PROPERTIES	GLAZING	
ADDRESS_HISTORY		This dataset will be used to figure out in which SOA the household lived in each year and when they moved.
ADDRESS_HISTORY	PREVADD	
ADDRESS_HISTORY	PREV_SOA2001	
ADDRESS_HISTORY	PREV_XUPRN	Needed to be able to link household to the property they were previously in.
ADDRESS_HISTORY	NILSID	
ADDRESS_HISTORY	SOURCE	

ADDRESS_HISTORY	CURRENT_FLAG	
ADDRESS_HISTORY	TYPE	
ADDRESS_HISTORY	SOA2001	
ADDRESS_HISTORY	XUPRN	Needed to be able to link household to the property they are in.
EVENTS		This data is used to link the births and deaths data to the households.
EVENTS	NILSID	
EVENTS	EVENT_TYPE_NAME	
EVENTS	EVENT_TYPE_CODE	
EVENTS	LINKID	
DEATHSSTATS		This data will be used to see how the composition of the household changes (i.e. if a member dies).
DEATHSSTATS	TYPEOFDT05	
DEATHSSTATS	GRODID	
DEATHSSTATS	BRTHMNTN	
DEATHSSTATS	BRTHYR	
DEATHSSTATS	DTHMNTN	
DEATHSSTATS	DTHYR	
DEATHSSTATS	AGE	
DEATHSSTATS	COUNTRYOFBIRTH	
DEATHSSTATS	COUNTRYOFUSUALRESIDENCE	
DEATHSSTATS	EMPSTAT	
DEATHSSTATS	EMPSTAT05	
DEATHSSTATS	ICD10CHAP	
DEATHSSTATS	MAINCAUSE	
DEATHSSTATS	MARITAL	
DEATHSSTATS	OCCCDE	
DEATHSSTATS	OUTSIDENI	
DEATHSSTATS	PLACEOFDEATHCODED	
DEATHSSTATS	SEX	
DEATHSSTATS	SOA_USRES	
DEATHSSTATS	SOCIALCLASS	
DEATHSSTATS	SOCIALCLASS01	
DEATHSSTATS	TYPEOFDT	
DEATHSSTATS	TYPEOFDT05	
BIRTHSSTATS		This data will be used to see how the composition of the household changes (i.e. if a member is born).
BIRTHSSTATS	GROBID	

BIRTHSSTATS	REGCOUN	
BIRTHSSTATS	OCCYR	
BIRTHSSTATS	OCCMONTH	
BIRTHSSTATS	OUTSIDENI	
BIRTHSSTATS	SEX	
BIRTHSSTATS	FAGE	
BIRTHSSTATS	MAGE	
BIRTHSSTATS	MARSTATOFPARENTS	
BIRTHSSTATS	DUROFMARR	
BIRTHSSTATS	PREVMARR	
BIRTHSSTATS	SOCIALCL	
BIRTHSSTATS	SOCIALCL01	
BIRTHSSTATS	FEMPSTAT	
BIRTHSSTATS	MEMPSTAT	
BIRTHSSTATS	EMPSTAT	
BIRTHSSTATS	EMPSTATNEW	
BIRTHSSTATS	MEMPSTATNEW	
BIRTHSSTATS	FEMPTSTATNEW	
BIRTHSSTATS	MOCCCODE	
BIRTHSSTATS	FOCCCODE	
BIRTHSSTATS	MSOCCCLASS	
BIRTHSSTATS	FSOCCCLASS	
BIRTHSSTATS	SOA2001	
BIRTHSSTATS	XUPRN	Needed to link the properties data.

Derived Variables

You may need to have new variables included in your project dataset which can be generated by the NILS-RSU using existing NILS/NIMS data (including confidential or restricted variables). Please list the input variables required for the derivation and, where possible, show how the new variable will be derived.

The data will be set up yearly (i.e. from 2001 to 2012). Household characteristics that do not change would be the same across each year (and others might be assumed to be constant) but it will have the property that the household lived in each year (this will change if the family moved house – which comes from the ADDRESS_HISTORY data) and the SOA where it is located.

I will want to calculate area characteristic variables (based on the census data) at SOA level for each year (e.g. the percentage of Catholics). Setting it up in this way will also allow the SOAs to be linked external area-based data.

Additional Information

Please outline any requirements you may have in relation to attaching external data (other than area-based data at SOA level or above) to your project dataset.

I will want to estimate a rough income measure for each households – this will be done by matching employment classifications and person characteristics to similar individuals that are in the ONS Labour Force Surveys (this is available from the UK Data Archive).

All other external data to be attached will be area-based at SOA level or above, this includes:

- House price trends – this will come from old property listings data
- Crime data – from the NINIS website and CAIN website
- Other area attributes – from the NINIS website

SECTION C1 LINKING INDIVIDUAL LEVEL DATA

A project that involves the linking of individual level data to the NILS is called a Distinct Linkage Project (DLP). DLPs require further steps to be undertaken to comply with legislation and safeguard the confidentiality of the data.

Description of Individual Level Data

Please describe the individual level data and provide information on the Data Custodian and how the data was collected.

Requirement for Individual Level Data

Please provide evidence of the research need for individual level data to be attached to the NILS and the additional research benefits that could be obtained.

Feasibility of Linkage

Please describe how you have investigated the feasibility of the linkage (e.g. the coverage of the Health & Care Number), the quality of the data and any experience in using the new data.

One-Way Encryption Methodology

The method of linking data using one-way encryption is described in Annex 1. Please confirm that it is appropriate for the data to be linked in this way. If the linkage requires any deviation from this methodology please describe it in detail below.

Legal Basis

Please provide details of any discussions regarding the legal framework for the data and the support of the Data Custodian.

- I have provided a letter of support from the Data Custodian.**
- I understand I will require ethical approval from ORECNI for a database modification.**
- I will adhere to the DHSSPS Privacy Advisory Committee Code of Practice of Protecting the Confidentiality of Service User Information.**
- I agree to keep NISRA informed of progress and will assist the NILS team in drawing up a Data Transfer Agreement between NISRA and the data supplier.**

Signature:

(electronic signatures accepted here)

Date:

Annex 1 - One-Way Encryption Process – for Distinct Linkage Projects

As part of the Data Transfer Agreement each Distinct Linkage Project requires encryption of the unique identification field (Health and Care number).

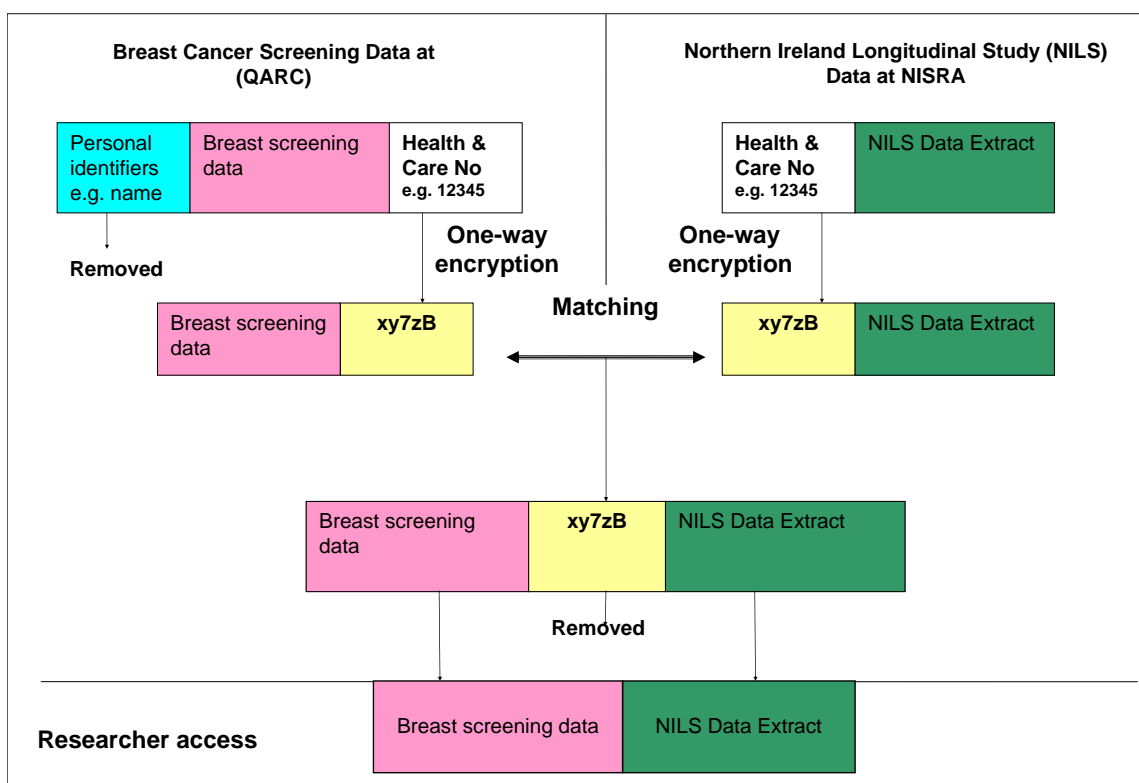
The NILS encryption software uses the Advanced Encryption Standard (AES) 256 algorithm for a one way encryption of the unique identifier and this will be carried out on both datafiles in situ. The software is designed to read in a comma separated text file with column headings where the first 10 characters are the unique identifier. It requires 2 distinct text strings, one is the password which forms the basis for the encryption key and the other is additional text to enhance the security of the encryption. Both these text strings will only be known by the external data provider so that the process cannot be replicated by NILS staff.

The resulting file contains the encrypted unique identifier plus the remaining variables (without the unique identifier) and is sorted in a random order to ensure that the encryption cannot be reversed on record order.

This file is then encrypted using 256 bit AES, for example using Winzip (version 9 or greater), for transport to the NILS secure setting where the NILS data extract undergoes the same process with the same password & key (again input by external data provider).

NILS staff then check both files to ensure fields are as specified in the Data Transfer Agreement. The 2 data files are then matched and merged using the encrypted unique identifier present on both files. This encrypted identifier is then removed from the final datafile. This finalised file is then made available to the approved Researcher(s) in the safe setting for analysis.

Diagram 1: Formation of a dataset for a NILS Distinct Linkage Project



A checklist, with signatures, is used by the relevant parties at each stage of data encryption.

REFERENCES

- Bayer, P., Ferreira, F. & McMillan, R., 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, 115(4), p.588-638.
- Bayer, P., McMillan, R., Murphy, A. & Timmins, C., 2011. "A Dynamic Model of Demand for Houses and Neighborhoods." *Mimeo*
- Barr, P.J. & Shuttleworth, I., 2012. "Reporting address changes by migrants: The accuracy and timeliness of reports via health card registers." *Health & Place*
- Cutler, D. M. & Glaeser, E. L., 1997. "Are ghettos good or bad?" *The Quarterly Journal of Economics*, 112(3):827-872.
- Cutler, D. M., Glaeser, E. L., & Vigdor, J. L., 2008. "When are ghettos bad? Lessons from immigrant segregation in the United States." *Journal of Urban Economics*, 63(3):759-774.
- Deloitte, 2007. "Research into the financial cost of the Northern Ireland divide."
- Echenique, F. & Fryer, R. G. J., 2007. "A measure of segregation based on social interactions." *Quarterly Journal of Economics*, 72(2):441-486.
- LPS, 2014. "Northern Ireland Residential Property Price Index October - December 2013." Land and Property Services
- Mueller, H., Rohner, D. & Schoenholzer, D., 2013. "Tectonic Boundaries and Strongholds: The Religious Geography of Violence in Northern Ireland," Cahiers de Recherches Economiques du Département d'Econométrie et d'Economie politique (DEEP) 13.04, Université de Lausanne, Faculté des HEC, DEEP.
- Office of the First Minister and Deputy First Minister, 2005. "A shared future: Improving relations in Northern Ireland."
- Office of the First Minister and Deputy First Minister, 2010. "Programme for cohesion, sharing and integration."
- ONS, 2012. "Measuring National Well-being, Where we Live, 2012." Office of National Statistics
- O'Reilly, D. & Stevenson M., 2003. "Mental health in Northern Ireland: Have the Troubles made it worse?" *Journal of Epidemiology and Community Health*, 57: 488-492
- Shuttleworth, I. & Barr, P., 2011. "Who reports address changes through the healthcare system? Describing the characteristics of laggards and non-reporters using the Northern Ireland Longitudinal Study". *Population Trends* doi: 10.1057/pt.2011.11
- Shuttleworth, I. & Lloyd, C., 2009, 'Are Northern Ireland's communities dividing? Evidence from geographically consistent Census of Population data, 1971-2001' *Environment and Planning A*, vol 41, no. 1, pp. 213-229., <http://dx.doi.org/10.1068/a40163>
- Tiebout, C., 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy*, 64(5): 416-24.
- The Marmot Review, 2010. "Fair Society, Healthy Lives: The Marmot Review strategic review of health inequalities in England post-2010."